APPLICABILITY OF NEURAL NETWORKS FOR IMAGE CLASSIFICATION ON OBJECT DETECTION IN MOBILE MAPPING 3D POINT CLOUDS

J. Wolf*, R. Richter, S. Discher, J. Döllner

Hasso Plattner Institute, Faculty of Digital Engineering, University of Potsdam, Germany
(johannes.wolf, rico.richter, soeren.discher, juergen.doellner)@hpi.de

KEY WORDS: Mobile Mapping 3D Point Cloud, Semantic Classification, Image Segmentation, Artificial Neural Network, Deep Learning, Object Detection

ABSTRACT:

In this work, we present an approach that uses an established image recognition convolutional neural network for the semantic classification of two-dimensional objects found in mobile mapping 3D point cloud scans of road environments, namely manhole covers and road markings. We show that the approach is capable of classifying these objects and that it can efficiently be applied on large datasets. Top-down view images from the point cloud are rendered and classified by a U-Net implementation. The results are integrated into the point cloud by setting an additional semantic attribute. Shape files can be computed from the classified points.

1. INTRODUCTION

Point clouds are widely used for storing geospatial information. They have proven to be a valuable data source for analyses as they are easy to handle and hold great detail of the captured environment (Vosselman et al., 2004). Technically, they are stored as an unordered collection of measurement points each featuring three-dimensional coordinates and additional attributes, e. g., intensity values when being measured via LIDAR (Richter et al., 2013). The unordered and unstructured points of a point cloud usually require a semantic classification for successive usage (Niemeyer et al., 2012). Semantic classification is the process of assigning each object an additional attribute that describes the type of the object, such as "Car", "Lamp post" or "Traffic sign – turn right". Once individual objects and their semantic classes have been identified, they can be used for, e. g., road cadastre creation or renewal (Caroti et al., 2005), clearance area checks (Mikrut et al., 2016), and 3D modeling (Vosselman, 2003). Typical semantic classes enable a basic distinction between ground, vegetation, and buildings, but many additional and very detailed classes might be needed for individual use cases, such as cars, road markings, traffic signs or curbstones (Pu et al., 2011). Figure 1 shows a point cloud for which semantic classes have been determined and are highlighted by assigning each semantic class an individual color.

In this work, we focus on a semantic classification using convolutional neural networks for visual recognition in images. In the past, many techniques for the automated analysis of images have been made and popular frameworks have been developed (Pulli et al., 2012). We show that those can also be used for the classification of point clouds in certain use cases. Some objects in road environments do not extend in three dimensions or have one dimension hidden below ground. Road markings and manhole covers are located flat on the ground and their height, if at all, differs only in a very small extent compared to their environment. These objects are predestined for visual recognition as they can easily be represented in two-dimensional images.

*Corresponding author

2. RELATED WORK

Point clouds are a valuable tool to automatically create 3D city models (Schwalbe et al., 2005) and landscape models for many different use cases in urban planning for local authorities, companies, or individuals (Vosselman et al., 2001). Cadastral data can be combined with point clouds to create interactive visualization tools for analysis and exploration (Aringer, Roschlaub, 2014). High density point information can be analyzed and creating large models gets possible even without a lot of human involvement (Richter, Döllner, 2013). Besides aerial captures of point clouds, mobile mapping techniques are widely used (Li, 1997). Mobile mapping scans can be used to, e. g., automatically extract road networks, or to analyze road surfaces (Jaakkola et al., 2008), as well as for the reconstruction of building facades.

For many use cases the automated analysis of point clouds is a mandatory preparation. Semantic classification can be performed by two fundamentally different approaches: Semantic per-point surface category information can be derived by analyzing a point cloud's topology (Chen et al., 2017) or by applying deep learning concepts (Boulch et al., 2017).

Traditionally, explicit rules are defined to distinguish semantic classes by geometric attributes (Grilli et al., 2017). Point clouds can be segmented into local groups of points with, e. g., similar surface directions (Rabbani et al., 2006). Each of these segments can then be analyzed with regard to their size and orientation. Large, vertical surfaces can then, for example, be identified as building facades, whereas groups of points whose corresponding surface normals are pointing into many different directions are usually part of vegetation (Wolf et al., 2019). An alternative approach uses machine learning techniques to identify the semantic classes of objects by using previously trained neural networks (Zhou, Tuzel, 2018). Such networks use an already classified dataset for training and learn to predict the semantic class for individual points or groups of points in new, unknown datasets. In recent years, using the internal structure of the point clouds themselves has become increasingly popular, as exemplified by PointNet and similar networks (Qi et al., 2017). However, they often focus on small datasets of separated objects and

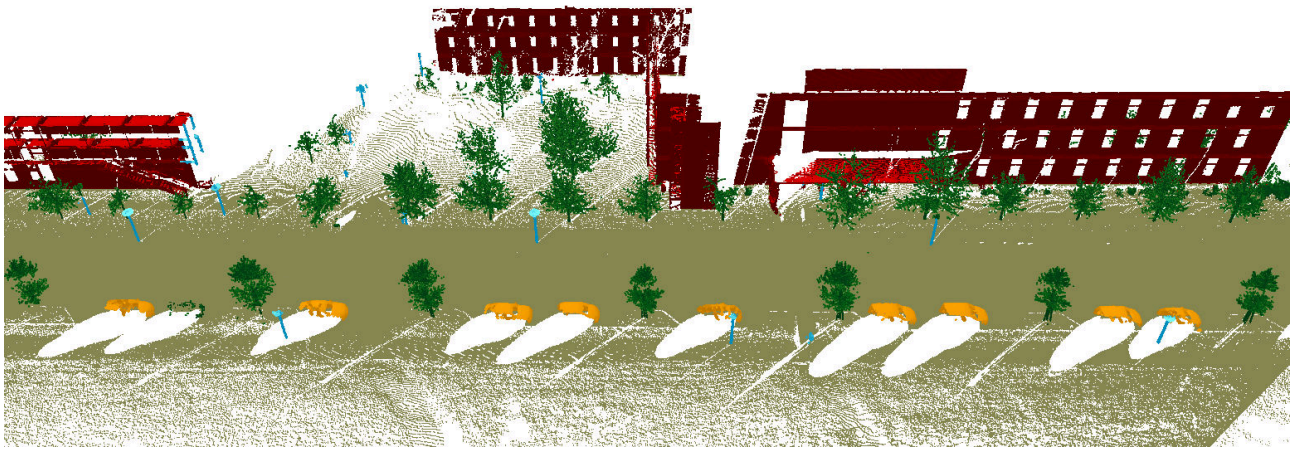


Figure 1. Point cloud colored based on semantic class: Ground (brown), vegetation (green), buildings (red), vehicles (orange), pole-like structures (blue).

require large training datasets.

Detecting objects in images is a well known research area that attracts great interest since many years. Viola et al. present an often cited image object detection algorithm which can be used to detect, e. g., faces in images (Viola et al., 2001). U-Net, originally developed in a medical context, is nowadays widely used for image segmentation (Ronneberger et al., 2015). It enables the automated detection of specific areas in images, such as cancer cells but also roads in aerial images (Zhang et al., 2018).

3. DATASETS

The datasets used in this work are mobile mapping scans from three different cities in Germany. They vary with respect to point densities and the number of cars, pedestrians, and other objects blocking the view. However, the trained network can detect objects in all datasets with similar accuracy. Different areas of the datasets were used for training and evaluation.

A typical street from a dataset is shown in Figure 2. The point cloud shown consists of 29 000 000 points and covers about 670 meters road with several crossings.

To create a training set for a U-Net-based neural network that is able to detect manhole covers, 300 images with manhole covers have been manually classified to create positive training data. Some images contain multiple instances. Several thousand road markings in 600 images have been marked as well. The training data set was completed with more than 16 000 images containing no manhole covers or road markings. Data augmentation, in this case rotation and mirroring, was used to enlarge the dataset even further.

4. CONCEPT AND IMPLEMENTATION

Our approach uses the abilities of image object detection algorithms to automatically classify certain objects in point clouds. We are focusing on manhole covers and road markings because they are clearly visible in a top-down view of the point cloud data. In Figure 3 the detected structures are highlighted.

We implemented a pipeline concept capable of automatically rendering large datasets, detecting objects in the created images, and mapping the results back into the original point clouds. Both object types were trained and detected separately.

First, all input point clouds are filtered as described in Section 4.1. A renderer then creates square images of these filtered point clouds as described in Section 4.2.

The rendered images are then classified with the previously trained neural networks and the results are integrated back into the point cloud, which is described in Section 4.3. Section 4.4 describes the creation of shapes for the individual objects.

4.1 Point cloud preprocessing

We aim to detect objects in billions of points of whole cities, so data reduction is an important aspect. To detect road markings and manhole covers, only the road itself along the captured path is required. During the scan of the point cloud data, a trajectory is captured, which describes the exact path of the measuring vehicle. The point cloud of interest can now be cropped along this trajectory, e. g., 10 meters to its left and right. If the recording path should not be available, relevant areas can also be filtered by analyzing the local point density, because regions close to the measuring vehicle have a much higher density than areas further away. In the remaining data, outliers are removed by simple outlier filtering. This is done to remove noise within the data that might affect the top-down rendering of the point cloud. All points with less than, e. g., five neighboring points within a distance of 0.5 meters can be marked as outliers. The approach can be sped up by using a heuristic search based on a spatial data structure such as a three-dimensional grid in which all points are placed. All points in cells which hold less than a certain number of points can be marked as outliers. For the specific use case of this work such a heuristic approach is sufficient, because the objects of interest are all located in dense areas of the point cloud.

A ground detection step identifies ground points based on their relative height and orientation (Meng et al., 2009). Higher points can be removed, so points representing buildings or high vegetation will not be analyzed. The algorithm divides the area that is to be analyzed into a regular two-dimensional grid. For each grid cell, the lowest of all z-values of the points falling into this cell, is stored. This represents a simplified terrain model. After the grid has been initialized, scan lines are used to find all ground points of the point cloud. These scan lines move axis-aligned in positive and negative direction as well as diagonally through the grid. The algorithm takes into account, which slope is determined in the different scanning directions and how the

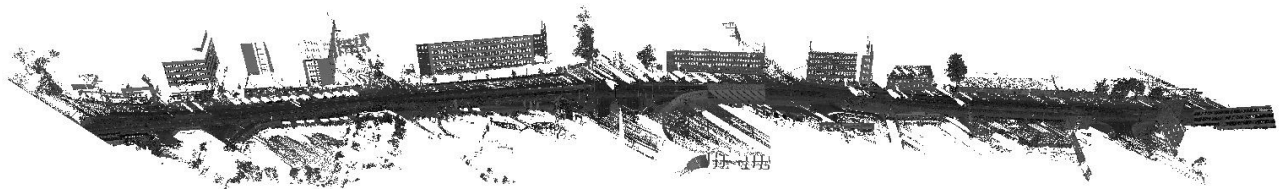


Figure 2. Point cloud of a street used as input for the object detection.

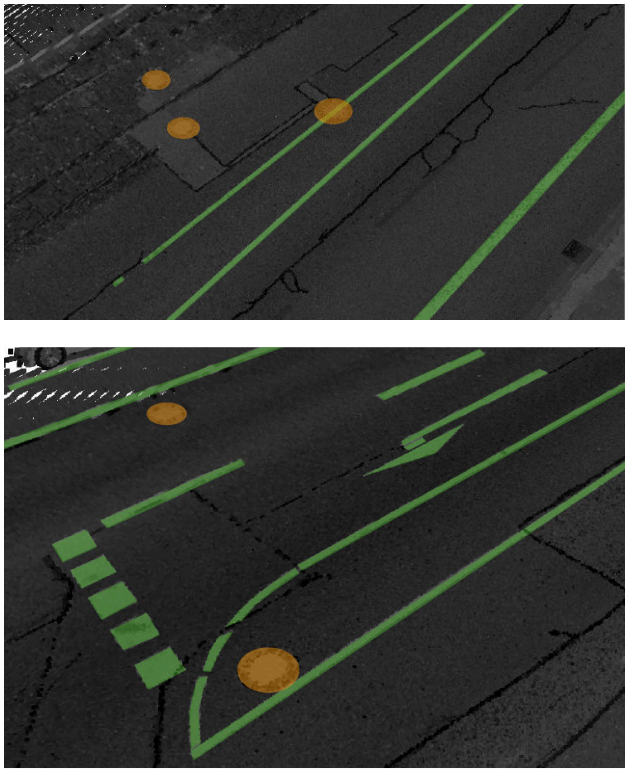


Figure 3. Point clouds with detected manhole covers (orange) and road markings (green) represented as shapes rendered on top of the point cloud.

elevation differs between points and the minimum elevation in their local neighborhood. For each scan line, potential ground points are determined separately. Following that, a majority voting is used to classify points as ground.

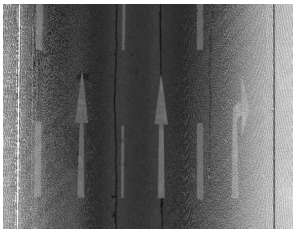
The remaining point cloud only consists of ground points along the measuring vehicle's trajectory without outliers. Following this preprocessing step, point clouds of our test dataset have on average about 60% of their original points left.

4.2 Image rendering

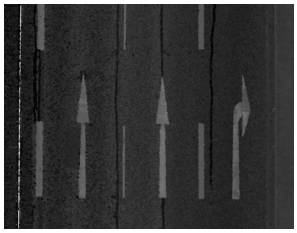
For the region of interest, images of the point cloud can now be rendered in a top-down view and are subsequently processed by the neural network.

Our renderer uses a point cloud as input and generates a series of images of 128x128 pixels in orthogonal projection, as shown in Figure 5. The visualized area has a size of about 4.5x4.5 meters. The positions of the images are selected in such a way that the complete area of the previously filtered point cloud is covered and that the images are overlapping. Each image contains a channel storing the intensity value of the point visible in each individual pixel as well as a channel storing the ID of the point that was rendered in this position. The latter is needed for projecting the classification results back into the point cloud.

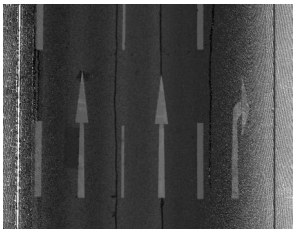
For best results, the points from the point cloud are rendered as paraboloids, of which only the tips are visible in dense areas. Rendering with different primitives is shown in Figure 4. Using paraboloids will fill more pixels in areas with lower density to avoid holes in the resulting image, while preserving sharp edges of individual structures, as shown in Figure 4g.



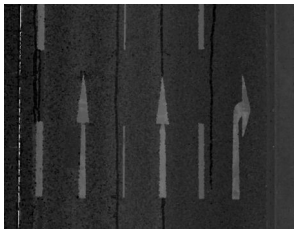
(a) Small points (2 pixels).



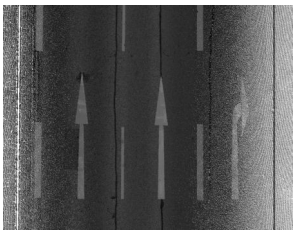
(b) Large points (6 pixels).



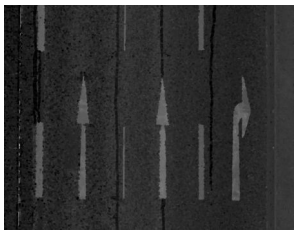
(c) Small circles (1x1 cm).



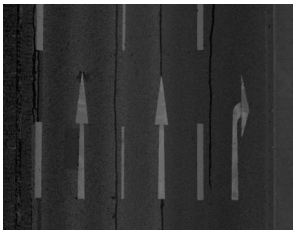
(d) Large circles (3x3 cm).



(e) Small quads (1x1 cm).



(f) Large quads (3x3 cm).



(g) Paraboloids.

Figure 4. Point cloud rendered with different primitives.

4.3 Classification

The rendered images are used as input for the previously trained neural networks. The result is an output mask for each input im-

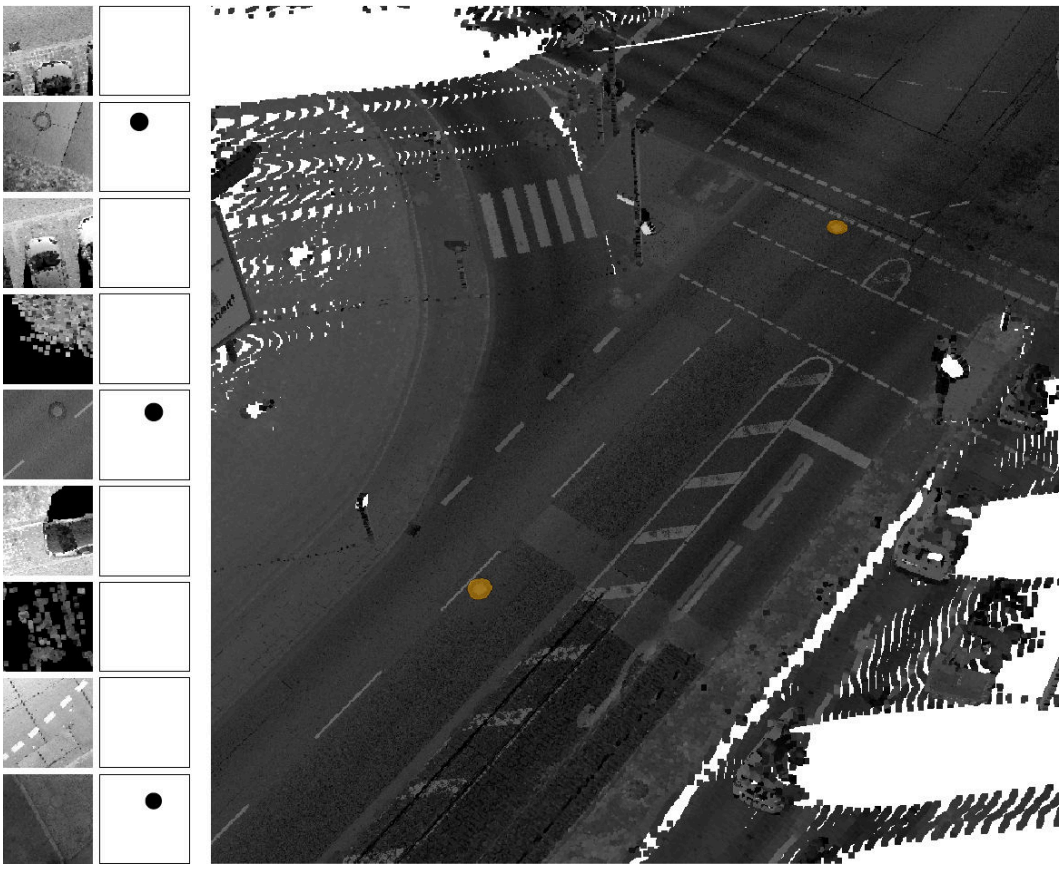


Figure 5. Left: Examples of rendered images and manually trained segmentation results for manhole covers. Right: Detected manhole covers projected back into the original point cloud.

age in which black pixels mark a detected object of the current type and white pixels everything else, similar to the training data shown on the left in Figure 5. After U-Net determined which pixels in an image belong to manhole covers, the associated points can be identified via the point ID layer. These points will get assigned an additional attribute referencing their semantic class. The number of points in the region covered by an image is often higher than the number of pixels in that image. Several points of the point cloud are therefore occluded and their IDs cannot be determined from the segmented images. Thus, points in the direct neighborhood of identified points will also receive the attribute value in a postprocessing step.

4.4 Generating shapes

For each group of adjacent points of a certain semantic class a shape can be created (ESRI, 1998). A convex hull or the best fitting rectangle is spanned around the points, resulting in shapes for road markings and manhole covers as shown in Figure 3. The resulting files can be used in various GIS applications for subsequent tasks.

5. EVALUATION

The evaluation dataset was a set of mobile mapping point clouds, captured with the same hardware setup in different areas of the cities also used in the training data. The neural networks were each trained for 5 hours on an Nvidia GeForce 1080 Ti. In an evaluation, 94% of the manhole covers and 91% of the road markings in the evaluation dataset were correctly identified and throughput of 7.5 million points per minute was reached. This number corresponds to approximately 300 meters of captured road. The bottleneck of the analysis are the I/O operations when rendering images of the point cloud and writing them to storage—an in-memory solution would highly increase the processing speed. Unidentified manhole covers usually had a low contrast to the street which could be an explanation why they have not been detected by the neural network. Road markings in areas with overall high intensity values have often not been found, especially in and around wet areas on the road.

6. CONCLUSION AND FUTURE WORK

We have shown that an established neural network for image segmentation can be used in the classification of flat objects such as manhole covers and road markings in mobile mapping 3D point clouds. By choosing an appropriate rendering technique, detailed images of the captured ground are created which can then be used as input for a neural network. It is possible to map the identified objects back into the point cloud as well as to create shape files which can be used in GIS applications.

We expect a similar approach to work on three-dimensional objects such as cars or traffic signs as well, but a number of modifications would be required that will be tested in future work. Multiple images can be taken from different angles around previously determined segments in a horizontal projection instead of a top-down view, e. g., four images from each side of the segment's bounding box. An additional depth channel can be added to the rendered images which includes information about the third dimension and allows for easy foreground/background

separation. Such an approach could then classify many additional objects via image segmentation that are not yet covered by the current analysis.

A combined network for multiple semantic classes would increase the processing speed compared to applying them one after another. Analyzing rendered images in memory without writing them to the hard drive would again increase the number of points that can be analyzed in the same amount of time.

REFERENCES

Aringer, K., Roschlaub, R., 2014. Bavarian 3d building model and update concept based on lidar, image matching and cadastre information. *Innovations in 3D Geo-Information Sciences*, Springer, 143–157.

Boulch, A., Saux, B. L., Audebert, N., 2017. Unstructured point cloud semantic labeling using deep segmentation networks. *Proceedings of 3DOR*, 2, 1.

Caroti, G., Piemonte, A., Pucci, B., 2005. Terrestrial laser scanning as road's cadastre revision and integration support. *ISPRS Workshop Italy-Canada 2005 3D Digital Imaging and Modeling: Applications of Heritage, Industry*, 1, CIRGEO, 1–3.

Chen, D., Wang, R., Peethambaran, J., 2017. Topologically aware building rooftop reconstruction from airborne laser scanning point clouds. *IEEE TGRS*, 55(12), 7032–7052.

ESRI, 1998. Esri shapefile technical description. https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf. Accessed: 2017-03-16.

Grilli, E., Menna, F., Remondino, F., 2017. A Review of Point Clouds Segmentation and Classification Algorithms. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 339.

Jaakkola, A., Hyyppä, J., Hyyppä, H., Kukko, A., 2008. Retrieval algorithms for road surface modelling using laser-based mobile mapping. *Sensors*, 8, 5238–5249.

Li, R., 1997. Mobile mapping: An emerging technology for spatial data acquisition. *Photogrammetric Engineering and Remote Sensing*, 63(9), 1085–1092.

Meng, X., Wang, L., Silván-Cárdenas, J. L., Currit, N., 2009. A multi-directional Ground Filtering Algorithm for Airborne LIDAR. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(1), 117–124.

Mikrut, S., Kohut, P., Pyka, K., Tokarczyk, R., Barszcz, T., Uhl, T., 2016. Mobile Laser Scanning Systems for Measuring the Clearance Gauge of Railways: State of Play, Testing and Outlook. *Sensors*, 16(5), 683.

Niemeyer, J., Rottensteiner, F., Soergel, U., 2012. Conditional Random Fields for LiDAR Point Cloud Classification in Complex Urban Areas. *ISPRS annals of the photogrammetry, remote sensing and spatial information sciences*, 1(3), 263–268.

Pu, S., Rutzinger, M., Vosselman, G., Elberink, S. O., 2011. Recognizing Basic Structures from Mobile Laser Scanning Data for Road Inventory Studies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(6), 28–39.

Pulli, K., Baksheev, A., Kornyakov, K., Eruhimov, V., 2012. Real-time computer vision with OpenCV. *Communications of the ACM*, 55(6), 61–69.

Qi, C. R., Su, H., Mo, K., Guibas, L. J., 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 652–660.

Rabbani, T., Van Den Heuvel, F., Vosselmann, G., 2006. Segmentation of Point Clouds Using Smoothness Constraint. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(5), 248–253.

Richter, R., Behrens, M., Döllner, J., 2013. Object Class Segmentation of Massive 3D Point Clouds of Urban Areas Using Point Cloud Topology. *International Journal of Remote Sensing*, 34(23), 8408–8424.

Richter, R., Döllner, J., 2013. Concepts and techniques for integration, analysis and visualization of massive 3D point clouds. *Computers, Environment and Urban Systems*, 45, 114–124.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, Springer, 234–241.

Schwalbe, E., Maas, H.-G., Seidel, F., 2005. 3D building model generation from airborne laser scanner data using 2D GIS data and orthogonal point cloud projections. *Proceedings of ISPRS WG III/3, III/4*, 3, 12–14.

Viola, P., Jones, M. et al., 2001. Rapid object detection using a boosted cascade of simple features. *CVPR (1)*, 1(511-518), 3.

Vosselman, G., 2003. 3D Reconstruction of Roads and Trees for City Modelling. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. Dresden, Germany*, 34, 3.

Vosselman, G., Dijkman, E., Reconstruction, K. W. B., Altimetry, L., Transform, H., 2001. 3D building model reconstruction from point clouds and ground plans. *Int. Arch. of Photogrammetry and Remote Sensing*, XXXIV, Part 3/W4, 37–43.

Vosselman, G., Gorte, B. G., Sithole, G., Rabbani, T., 2004. Recognising Structure in Laser Scanner Point Clouds. *International archives of photogrammetry, remote sensing and spatial information sciences*, 46(8), 33–38.

Wolf, J., Richter, R., Döllner, J., 2019. Techniques for automated classification and segregation of mobile mapping 3d point clouds. *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 201–208.

Zhang, Z., Liu, Q., Wang, Y., 2018. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 749–753.

Zhou, Y., Tuzel, O., 2018. Voxelnet: End-to-end learning for point cloud based 3d object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4490–4499.